JOURNAL OF BACTERIOLOGY, June 1995, p. 3361–3369 Vol. 177, No. 12
0021-9193/95/$04.00+0

MINIREVIEW

Physical Mapping of Bacterial Genomes

MICHAEL FONSTEIN AND ROBERT HASELKORN*

Department of Molecular Genetics and Cell Biology, The University of Chicago, Chicago, Illinois 60637

INTRODUCTION

The elucidation of gene order in chromosomes, influencing both linkage and coregulation, is at the heart of genetic analysis. Indeed, the characterization of any mutation or cloned gene traditionally includes the determination of its position on the chromosomal map. Further, questions about the structural maintenance of the chromosome itself can be approached by comparative physical and genetic mapping of chromosomes. However, viewing the genome in one dimension may not be sufficient to understand genome organization comprehensively, if the object of evolutionary conservation is its three-dimensional structure.

Beginning in the early 1980s, several experimental approaches made it possible to map megabase-size DNAs: pulsed-field gel electrophoresis (PFGE)-related methods, gene encyclopedia construction, and total DNA sequencing. The ideal targets for these physical approaches are bacterial genomes, with sizes ranging from 600 kb for *Mycoplasma genitalium* (22) to 9.5 Mb for *Myxococcus xanthus* (14). Aligning genes to physical maps generated by these approaches has surpassed traditional techniques for constructing bacterial genetic maps.

Many chromosomal maps of different bacteria have been produced by traditional genetic methods. Some of these maps have hundreds of genetic markers: there are 1,717 on the last version of the *Escherichia coli* map, of which 300 were located by physical mapping (71). Purely genetic maps have a few serious disadvantages. First, their construction is labor-intensive. Also, because genetic maps provide distances measured in units derived from recombination frequencies, they are sensitive to recombination hot spots. The physical distance equivalent to 1% of the *E. coli* linkage map varies from 33 to 60 kb (80). The genetically linear chromosomal map of *Caulobacter crescentus* turned out to be physically circular, with proximal (unlinked) markers being only 400 kb apart. This genetic separation is ascribed to a recombination hot spot in the Ter region (31).

The basic questions mentioned above, plus the availability of several convenient mapping techniques, have resulted in the construction of physical maps of many bacterial genomes. More than 90 studies dealing with physical mapping of bacterial genomes were published in the last 7 years. The five purely genetic maps produced during the same period is a fair reflection of their relative difficulty. In this minireview, we summarize the principal methods used for map construction, salient features of the maps produced, and applications of the gene encyclopedias for questions of genome evolution, rapid mapping of cloned genes, and global studies of gene expression. We also refer readers to a recent article by Cole and Saint Girons (21), published as this minireview was being completed. They cover much of the same ground and include tabular compilations of methods used and genome sizes. Their article should also be consulted for speculations on the evolution of genome size and the conservation of genome architecture.

* Corresponding author. Mailing address: Department of Molecular Genetics and Cell Biology, The University of Chicago, 920 East 58 St., Chicago, IL 60637. Phone: (312) 702-1081. Fax: (312) 702-3172. Electronic mail address: fons@midway.uchicago.edu or rh01@midway.uchicago.edu.

MAPPING STRATEGIES

PFGE-generated maps. In conventional gel electrophoresis, all DNA fragments larger than a certain size migrate with the same speed. This size depends on the agarose concentration and is about 50 kb in 0.4% agarose gels. To start moving, DNA fragments have to be properly oriented along the force lines of the electric field. The time required for this orientation is proportional to the size of the DNA fragments over a broad range of sizes. Large molecules can be effectively separated in a constantly reorienting electric field on the basis of this size-dependent time required for reorientation.

A device based on this principle, together with its mapping applications, appeared in a revolutionary article introducing PFGE (76). The originally nonhomogeneous electric field in this early version of PFGE has been changed to a uniform electric field in numerous improved versions, employing separately regulated point electrodes and different angles between switching electric fields, rotating the gels, or switching the direction in the thickness of the gel. The upper limit of reproducible size resolution has been increased to 10 Mb, far beyond the needs of physical mapping in prokaryotes.

The problem of generating high-molecular-weight DNA to be used as the substrate for separation in PFGE was solved by embedding whole bacterial cells or protoplasts in low-melting-point agarose with subsequent lysis. After detergent and proteolytic treatments, cell components are simply washed out, leaving intact chromosomes immobilized in agarose. The use of pipettable agarose beads or melting DNA samples after digestion (for restriction fragments smaller than 1,000 kb) makes the loading of tiny agarose blocks unnecessary. If conditions for the removal of cell walls from target cells can be found, subsequent preparation of DNA samples from most sources is routine and highly reproducible.

Physical genome analyses require restriction enzymes with recognition sequences (RSs) that can split bacterial genomes into a limited (1 to 20) number of fragments. When the PFGE epoch began, we had only two restriction enzymes with eight-base sites, *Not*I (GCGGCCGC) and *Sfi*I (GGCCnnnnGGCC). We now have nine others, both AT-rich (*Swa*I [ATTTAAAT], *Pac*I [TTAATTAA], and *Pme*I [GTTTAAAC]) and GC-rich (*Sgr*AI [CACCGGCG], *Sse*8387I [CCTGCAGG], *Srf*I [GC CCGGGC], *Sgf*I [GCGATCGC], *Fse*I [GGCCGGCC], and *Asc*I [GGCGCGCC]), listed in the Rebase database of R.

Roberts. Reduced representation of certain oligonucleotides in bacterial genomes and simply different GC contents of target DNAs and RSs increase the number of restriction enzymes that can be used as "rarely cutting," adding many enzymes with six-base RSs. A full list of such enzymes with respect to the GC composition of bacterial chromosomes has been published elsewhere (57).

Another source of rarely cutting restriction enzymes are the proteins that participate in intron processing. Having very large RSs of 18 bp or even 26 bp (54), they can provide unique cleavage points in bacterial genomes and can also be very useful markers for certain genetic elements. For instance, the 26-bp RS of *Ceu*I is conserved in the *rrl* gene (23s rRNA) of many eubacteria (54) and absent from the rest of the genome. The *Ceu*I restriction maps for several strains of *E. coli* and *Salmonella typhimurium* revealed almost absolute conservation of the *rrn* "skeleton" for these two genera. For more than a dozen physically mapped bacteria, *rrn* operons seem to conserve sets of RSs extremely rare in the rest of the genome.

Additional tools are provided by the introduction of rare cleavage sites to the genome. Tn*5* insertions can be conveniently localized on restriction fragments in PFGE because of the *Not*I site in its IS sequence (81). This approach, with numerous modifications, was successfully applied in at least a dozen mapping projects.

The lambda *cos* site (with a minimal functional size of 60 bp) can be delivered to specific sites by an integrating vector (87). The in vitro interaction between the *lac* operator and *lac* repressor can be used to protect the *Hae*II site in the *lac* operator from methylation; subsequent *Hae*II cleavage split the 160-kb *Not*I fragment of *E. coli* and entire yeast chromosomes uniquely where *lacO* was integrated (46).

Other methods depend only on the availability of sequencing information in the area to be cleaved. A model oligonucleotide able to form a stable DNA triple helix led to specific unique cleavage, because of the methylation protection of this region (83). The actual integration of such a sequence is not necessary because DNA stretches capable of forming such structures can be found in the sequences of most target organisms. By adding RecA to a similar system, one can use an oligonucleotide complementary to any target DNA (45).

The three elements, the ability to extract intact bacterial chromosomes, the ability to cut them into a limited number of pieces, and the ability to separate those pieces by PFGE are the basis of long-range physical bacterial genome mapping. Strategies designed to order the fragments observed with PFGE are discussed below.

All existing mapping procedures have weak points. Only by applying several procedures together can a reliable physical map be produced. The first physical-genetic map of a bacterial genome was constructed for *E. coli* EMG2 (80). Most of the approaches used in recent mapping projects were introduced in this article (80) and the following article (81). The use of linking probes and Tn insertions marked by a rare cutting site and establishment of the fragment order by probing partial digests with a unique probe were all used to order 22 *Not*I fragments on the 4.7-Mb chromosome of *E. coli*. However, the important conclusion—colinearity of the physical and genetic maps of bacteria—was not rigorously justified, because a considerable part of the fragment ordering was done according to the genetic positions of the probes used. A detailed practical comparison of the efficiencies of existing mapping strategies can be found elsewhere (64). All the approaches mentioned above plus two-dimensional (2-D) PFGE and hybridization fingerprinting of macrorestriction fragments (MRFs) were used to locate 25 *Spe*I and 33 *Ase*I MRFs in the 9.35-Mb genome of *Stigmatella auriantica* DW4/3.1 in this study.

Methods based on DNA hybridization. (i) Hybridization with eluted MRFs. The hybridization of PFGE-generated gel blots with eluted MRFs has been used widely to link fragments produced by two different enzymes. Repeating such steps can in principle expand the linked area until the map is completed. Problems with fragment ordering arise if the overlap region is significantly shifted toward one end of a probe or if a probe covers more than three MRFs. In addition, repeated elements can produce false positives, as in all approaches based on hybridization. In at least one study (*Neisseria gonorrhoeae* FA1090 [28]), linking with MRFs was unreliable because of repeats. Even the smallest MRF probe revealed numerous confusing bands. The ability of random PCR to reduce the complexity of a probe generated on a template of an eluted MRF, which avoids most problems due to repeats, is demonstrated in a partial *Sma*I and *Csp*I map of *Staphylococcus aureus* (69).

(ii) Unique probes. Another variation is the use of unique cloned fragments to staple MRFs generated by several enzymes. One such probe will not link two fragments but only establish an overlap for MRFs generated by different enzymes. However, by using a sufficient number of probes, it is possible to assemble the whole genome map. Very often this is used in combination with partial digestion, where a unique probe for a particular MRF is used to visualize the products of partial digestion of the MRF (21). The sizes of adjacent fragments can then be calculated with data from a complete digest. The fragments used for probing may be mapped genes or simply random clones. Locating gene probes can also be done in the opposite order, by transformation of the MRFs into auxotrophic strains.

(iii) Linking probes. Adjacent MRFs can also be assigned by using linking probes, which are cloned DNA fragments harboring rare RSs. Hybridized against a PFGE-generated blot, such a fragment should reveal two adjacent MRFs. Selective schemes in which insertion of an antibiotic resistance marker into rare sites was used to find linking clones, together with the direct restriction screening of gene libraries, have been devised. To find linking probes in a gene library of *Synechococcus* sp. strain PCC 7002 (15), chromosomal DNA was digested with *Not*I, cohesive ends were labeled by polymerase reaction and used as a probe after *Rsa*I digestion, which removed most of the MRF. Modification of the use of linking probes has also been shown (42), where 26 *Sfi*I and most of the 70 *Not*I sites were mapped onto the 4,165-kb genome of *Bacillus subtilis* 168 (CU741). Rare sites were mutated in the linking clones and reintroduced into the chromosome. The disappearance of two fragments, indicating linkage, was detected in all these cases. Another top-down approach was used to link 25 *Ase*I and 14 *Xba*I fragments in the 3.7-Mb chromosome of *Rhodobacter capsulatus*. Eluted MRFs were blotted onto a dense-grid array of a cosmid library. Cosmids revealed by two different MRFs linked adjacent fragments (34). As with any method based on hybridization, the use of linking clones has a tendency to produce false positives, because of repeated sequences.

(iv) Hybridization fingerprinting. To reveal fragments that can link MRFs of interest, chromosomal digests generated with an "intermediately rare cutter" were probed with these MRFs. If two MRFs identified bands with similar mobilities, they were said to be linked (64).

Combinatorial methods. Methods based on combinatorial analyses of different multiple digests are not sensitive to repeats and can be applied to organisms carrying many repeated sequences. However, these methods are more dependent on

the accuracy of PFGE than other strategies and are not always able to link all MRFs into a unique structure. The most advanced among this group of methods is 2-D PFGE, applied to genomes ranging in size from 780 kb for *Mycoplasma mobile* ATCC 43663 (4) to 5.9 Mb for *Pseudomonas aeruginosa* PAO (73) and even 9.4 Mb for *Stigmatella auriantica* DW4/3.1 (64). Chromosomal fragments digested partially or completely and resolved by PFGE are subjected to full digestion with the same or another rarely cutting enzyme. The products of the second digestion are separated by PFGE in the direction perpendicular to the first PFGE run. MRFs of the second digest that are derived from the same fragment of the first digest are positioned at the same latitude. In principle, one 2-D gel is sufficient to align all MRFs generated by a pair of rarely cutting enzymes (by simple combinatorial analysis). Revealing the MRFs via end labeling instead of ethidium bromide staining significantly improves the image quality (73). However, this method is very sensitive to gel imperfection. A less sophisticated modification of this method is the redigestion of eluted MRFs, which looks less state of the art but is easier to set up than the 2-D gels.

In the case of a very limited number of fragments, simple comparison of the products of full and partial digests can produce a restriction map. Six *Hpa*I sites were positioned on the 1.74-kb chromosome of *Thermus thermophilis* this way (7). Unfortunately, this approach depends on the accuracy of size evaluation in the size range where this is least reliable. The positions of a few restriction sites (usually less than five for each enzyme) in the chromosomes of several bacteria were derived from such partial digests, for example *Thermococcus celer* (66). A similar combinatorial approach based on double digest analysis was used to place a few of the *Swa*I, *Pac*I, *Pme*I, and *Ceu*I sites onto the chromosome and megaplasmids of *Rhizobium meliloti* 1021 (40).

A cartridge carrying a RS was placed into 21 genetically mapped genes of *Bradyrhizobium japonicum* 110 to reveal overlapping MRFs by splitting them (48). An early *Not*I map of *B. subtilis* was generated via 54 genetically mapped integrated Tn*917* (86). Most of the physical maps for *Salmonella* spp. are also based on a genetically mapped Tn*10* collection (55). In addition to linking genetic and physical maps, Tn elements were used as random physical markers, helping to establish MRF overlaps by introducing sites for rarely cutting enzymes or simply as targets for blot hybridization. Recently, a special derivative of Tn*4001* (harboring 32 unique sites for rarely cutting enzymes) was used for the physical mapping of *Acholeplasma oculi* ISM1499 (85).

GENE ENCYCLOPEDIAS

In 1987, the advantages of another approach to the physical study of bacterial genomes, the so-called gene encyclopedia, were demonstrated in the article by Kohara et al. (44). We apply the term gene encyclopedia to a gene library, in which cloned DNA fragments are ordered in a sequence reconstructing the gene order of the chromosome. The term N-fold or N-hit gene library is used regarding the representivity of the library, where N is the number of genome equivalents calculated as the number of clones multiplied by the insert size and divided by the genome size. A 10-hit library is generally sufficient for a mapping project. A contig is an uninterrupted group of linked or ordered clones. A miniset is an ordered group of clones covering a genome of interest with minimal overlaps. A gap in a gene encyclopedia is an area of the genome not represented in the encyclopedia because of unclonability or statistical reasons. The latter is called a statistical gap.

Construction of a gene encyclopedia can be divided into two stages. The first stage is the construction of a representative gene library with the minimal number of rearranged clones. The second stage is assembly of the clones by detection of their overlaps. As in PFGE-related projects, there are two ways to detect overlaps. One is based on DNA hybridization, and the other is based on a comparison of restriction maps or patterns.

Library construction. Two major factors, stability of the cloned material and insert size, influence the choice of a cloning system. Yeast artificial chromosomes (YACs), λ vectors, P1 vectors, cosmids, and Ori F plasmids are now available as tools for the cloning of large DNA fragments. λ vectors, used in early projects of gene encyclopedia construction (*Saccharomyces cerevisiae* [67] and *E. coli* [44]) are believed to guarantee stability of the cloned material. However, seven gaps in Kohara's *E. coli* encyclopedia in λ had to be closed by using the cosmid pOU61 (43). Cosmid vectors, mainly harboring λ Ori (36), were used for the construction of many bacterial encyclopedias. These λ Ori cosmids have constant copy numbers, independent of the size of the insert, which greatly reduces segregation selection for deletion variants. YACs have the advantage of very large insert size. YAC4 was used for *B. subtilis* (2), known for unstable propagation of its cloned DNA in *E. coli*, and for *M. xanthus* (49). In the first work, 10% of the YACs were found to be rearranged by hybridization with genetically mapped genes and had to be removed from further study. P1-derived vectors, which accommodate 90-kb inserts, have been used successfully as supplementary systems in many eukaryotic projects (for example, reference 39). For the promising new group of Ori F cloning vectors (79), not enough experience has been accumulated yet. At present, λ Ori cosmids like Lorist 6 (36) appear to be among the most reliable and convenient tools for gene library construction and encyclopedia assembly.

Encyclopedia assembly. Ordering the clones of a gene library requires establishing the overlaps between them. An important parameter, the minimal detectable overlap (MDO), is the portion of the insert required for detection of linkage related to the whole insert. The difference in the number of clones necessary to assemble a certain number of contigs is 10-fold between cases of 20 and 80% MDOs (9). Estimates predict 10 gaps (rather, unfound links) in a sevenfold library in the case of 20% MDO, whereas 60 gaps will exist in the case of 50% MDO (25). Criteria for a good fingerprinting approach were formulated in this review. However, hybridization methods were unjustifiably excluded from consideration. One article (24) is devoted to the comparison of fingerprinting and hybridization strategies in clone alignments. Fingerprinting was found to be insensitive to repeats and to have a small (log) increase in mapping efforts with increase in the size of the project. For hybridization approaches, reduction of the size of the MDO compensates for linear increase of both probe and clone number caused by an increase in the genome size. These efforts can be further reduced by using different means of grouping the hybridized clones. A total of 36,000 colonies can be spotted on filters (22 by 22 cm) by a new robotics device and analyzed at once (59). Alternatively, plating 3X3X96 colonies per petri dish-size filter is easy and does not need expensive hardware (for example, reference 34).

Fingerprinting approaches. Purely "fingerprint assembly" was used in the initial stages of the *S. cerevisiae* mapping project (67). *Eco*RI plus *Hin*dIII restriction patterns of 5,000 λ clones were assigned to 680 contigs. With an average of eight restriction fragments per hybrid clone, the MDO for this study was >50%. More-precise measurements were used in genome mapping of *Caenorhabditis elegans* (23). Cosmid clones were

cleaved with *Hin*dIII, end labeled, and redigested with *Sau*3A. The resulting fragments were separated with sequencing gels, and only labeled fragments were visualized on X-ray films. The high accuracy of separation increased the reliability of established overlaps. Still, the MDO was about 50%. This stage of the genome project of *Caenorhabditis elegans* ended with 860 contigs ranging from 35 to 350 kb. Eventually, both the *S. cerevisiae* and *C. elegans* projects were finished successfully by using combined approaches.

The application of direct fingerprinting to prokaryotic projects can be illustrated with several examples. For *Desulfovibrio vulgaris*, 879 λ clones representing its 1,720-kb genome were combined in 87 contigs (26). Similar work with *E. coli* MG1655 produced 68 contigs (25). However, by using the same pattern comparison with sequencing gels, it was possible to assemble 1,000 cosmids of a *Mycobacterium leprae* library into only four large contigs (30). *Hin*fII fingerprinting of the cosmids from an *E. coli* 803 library produced 58 contigs, covering four of seven gaps in the Kohara library. More-directed search and a low-copy-number vector were used to close the remaining three gaps (43).

A total of 1,025 λ clones from an *E. coli* W3110 library were assembled in 70 contigs covering 94% of the genome by comparing restriction maps of the cloned DNA fragments (44). A criterion of having at least five common fragments in an overlap gives 10 to 25% MDO. As calculated in reference 25, this guarantees less than 1/1,000 false overlaps. "Cosmid walking" with another 2,344 clones produced the final version of the Kohara map with only seven minor gaps. Similar mapping of YAC clones corresponding to the 9,454-kb chromosome of *Myxococcus xanthus* (49) linked 409 YACs into 60 contigs.

So-called landmark analysis was used to group cosmids from a library of *Haloferax volcanii* DS2 around "semirare" sites (13). To recognize individual sites, cosmids were digested with a frequently cutting enzyme together with a semirare one. Two restriction fragments surrounding a semirare site can be uniquely recognized for most chromosomal locations. Fifty-nine contigs were further linked by cosmid walking and PFGE to produce seven contigs.

Linking by hybridization. Most of the prokaryotic encyclopedias were assembled by hybridization, an expansion of genome walking. If a cloned insert or its ends is used as a probe against a gene library, it will identify a group of overlapping clones. Since a representative number of clones (1 to 2,000 for a prokaryotic genome) can be plated in a grid array, it is possible to do several such hybridizations at a time. Selection of new probes among "negative" clones can be used to optimize the procedure, first applied to *Mycoplasma pneumoniae* (88). The 1.7-Mb genome of *Helicobacter pylori* was covered by 60 such steps grouped in five consequent sets (10). Three hundred steps (with large redundancy) were used in the complete mapping of the 3.7-Mb genome of *R. capsulatus* (33). A total of 764 probes were used for the 14-Mb genome of *Schizosaccharomyces pombe* (39).

Several elements are essential for successful cosmid alignment. One element is a preliminary grouping of clones using MRFs or YACs. This helps choose evenly dispersed probes for the initial step of walking. An improvement is to pool the probes in a 2-D scheme. Groups of 12 probes each, in which each cosmid is represented in two groups, were used as probes (39). This approach increases the informational content of hybridization patterns, reducing the number of hybridizations required. However, this approach is very sensitive to failures in experimental procedure.

Another key element in assembly is the size of the probe. The use of phage promoters in the vector, transcribing the ends of the inserts in vitro, has been replaced by primer extension using PCR (62). This can reduce the probe size from 5 to 15 kb to 500 bp, reducing the chance of hitting repeated elements, which causes ambiguities in assembly. On the other hand, whole cosmids were used as probes to reveal all possible repeats in the genome of *S. pombe* (39).

Oligonucleotide fingerprinting. In the latter work, a set of 11- and 12-nt oligonucleotides was also used as a supplement for cosmid grouping (39). A discussion of this technology can be found elsewhere (35). We can expect substantial improvements in this area as a by-product of the development of "sequencing by hybridization" with arrays of oligonucleotides.

Encyclopedias derived from genetic maps. A total of 310 clones of the Clark-Carbon collection selected from 2,200 ColE1 plasmids harboring 10- to 14-kb inserts of the *E. coli* genome were ordered by genetic complementation, hybridization with cloned genes, and protein overexpression and positioned according to the *E. coli* genetic map. A total of 518 other clones were mapped by hybridization with the Kohara miniset (65). YACs of an encyclopedia of *B. subtilis* were assigned by using the genetic map of this organism (2).

Summarizing, it becomes obvious that there is a wide choice of strategies for encyclopedia assembly in bacterial genome projects. In our view, a combination of hybridization using small probes with limited restriction mapping of the encyclopedia elements provides the most efficient way to assemble complete maps without ambiguities.

SEQUENCING PROJECTS

After a decade of scientific development, it has become apparent that the resolution of genome maps based on PFGE and encyclopedias is often insufficient. PFGE analysis and encyclopedia construction and alignment should be considered as important but preliminary stages on the way to deeper knowledge that can be achieved only by large-scale sequencing projects supplemented with a functional description of the genome.

The *E. coli* 1993 update of the compilation of DNA sequences (47) involves 2,353 kb (49.87% of the genome) of sequences. For *B. subtilis*, 650 kb was represented in nonredundant parts of 450 GenBank entries (37). Two to three hundred kilobases are sequenced for at least a dozen other microorganisms. These data are the result of random efforts of the scientific community.

Parallel to random sequencing, systematic sequencing projects have been started with several microorganisms. The most progress has been achieved in the *E. coli* genome project. Twenty percent of the genome between min 80 and 100 was sequenced in Madison, Wisconsin (for example, reference 6). An area between min 0 and 33 is being studied in several laboratories in Japan. About 140 kb of the *B. subtilis* genome has already been published as part of the work of the European sequencing consortium (37, 82). More than 220 kb of the chromosome of *Mycobacterium leprae* has been sequenced (89). Analysis of the first sequenced cosmid revealed 12 identifiable genes, representing 40% of the coding capacity of the sequenced region (41). Gene encyclopedias were used as the starting material in all these projects.

A less laborious version of a total genome project, a genome survey, was applied to *Mycoplasma genitalium* to provide markers for its physical mapping (68). Recognizable pieces of 97 proteins, 12 IS elements, and an *rrn* operon were found in shotgun sequencing of 508 random clones. The sampled elements cover 340 kb, that is, 59% of the 600-kb genome of this organism.

Genome sequencing projects are usually accompanied by systematic functional analyses. For example, 23 transcripts from vegetative and sporulating cells of *B. subtilis* were mapped for a large sequenced region, *spoVAF* to *serA* (3). For *E. coli*, transcripts generated under seven different conditions were located by using the Kohara miniset as the template for hybridization (18). This work was done with the ordered set of clones, without applying sequencing information. However, the use of PCR-generated templates, corresponding to open reading frames (based on sequences) increased the specificity of such mapping at least 10-fold (17).

CURRENT STATUS OF BACTERIAL MAPS

Physical maps are now available for at least 93 organisms. Eighty of the 93 were constructed by PFGE, and 13 were constructed via gene encyclopedias. Different numbers of independent physical maps of species in the following taxonomic groups have been done: *Streptomycetes*, 2; *Archae*, 7; *Neisseria*, 3; *Salmonella*, 4; *Rhizobium* and *Bradyrhizobium*, 2; *Borrelia*, 2; other *Spirochaetacea* spp., 5; *Campylobacter*, 5; cyanobacteria, 3; *Enterococcus*, 1; *Mycoplasma*, 10; *Lactococcus*, 2; *Myxococcus*, 2; *Stigmatella*, 1; *Streptococcus*, 3; *Haemophilis*, 3; *Clostridium*, 2; *Pseudomonas*, 3; *Rhodobacter*, 2; *Escherichia*, 8; *Bacillus*, 7; *Helicobacter*, 2; *Mycobacterium*, 2; *Bordetella*, 1; *Caulobacter*, 1; *Listeria*, 1; and *Acholeplasma*, 1.

Genome size, plasmids, and chromosome topology. Physical methods provided direct measurements of genome content for various microorganisms. The smallest genomes were found among *Mycoplasma* spp. (600 to 1,200 kb). Microorganisms with genome size around 2 Mb exist in various groups (many *Archae*, *Neisseria*, *Helicobacter*, *Haemophilus*, and *Streptococcus* spp.). *Rhodobacter*, *Salmonella*, *Escherichia*, *Bacillus*, and *Pseudomonas* spp. all have genomes around 4 to 5 Mb. The largest genomes were found in microorganisms capable of complicated differentiation, such as *Streptomyces* spp. (8 Mb), *Bradyrhizobium* spp. (8.7 Mb), *Myxococcus* spp. (9.5 Mb), and *Anabaena* spp. (6.5 Mb). Clustering of the genome sizes in a 1-2-4-8 Mb progression is consistent with old ideas of evolution via duplication or merging of different genomes.

On the other hand, genome size may vary drastically within a single taxonomic group (1 to 5 Mb for *Spirochaeta* spp.; 2.7 to 6.5 Mb for cyanobacteria). Events significantly changing genome size, such as 2-Mb deletions in *Streptomyces ambofaciens*, have been identified (50). The genome sizes of different strains of the same species, *S. ambofaciens*, can vary from 6.5 to 8 Mb.

More than one large replicon was found in several microorganisms. This is not an essential property of a distinct group, since closely related species can differ in the arrangement of their genomes. For example, the genome of *R. capsulatus* consists of a 3.7-Mb chromosome and often a 130-kb plasmid (34), but two chromosomes, 3.05 and 0.9 Mb, were found in the closely related *Rhodobacter sphaeroides* (84). The genome of *R. meliloti* is divided into 3.4-, 1.7-, and 1.4-Mb chromosomes (40), while the related *B. japonicum* has one 8.7-Mb chromosome (48). *Pseudomonas cepacia* harbors 3.4-, 2.5-, and 0.9-Mb chromosomes (16), while *P. aeruginosa* PAO has one 5.9-Mb chromosome (73).

According to the taste of the researchers, the 900-kb element of *R. sphaeroides* was called the second chromosome, while the two 1.7 and 1.4-Mb elements of *R. meliloti* were called plasmids. Most of these smaller elements carry essential genes, thus deserving the name chromosome. Even the 350-kb "plasmid" of *Leptospira interrogans* carries such genes (89). However, none of several hundred auxotrophic mutations of *H. volcanii* was complemented by DNA fragments from its three large plasmids, which comprise 1/3 of its genome (20).

The dogma concerning the circular structure of bacterial chromosomes was shaken after the discovery of linear chromosomes of *Borrelia burgdorferi* (32). Another example was added in 1993, when such chromosomes were found in *Streptomyces* spp. first for *Streptomyces lividans* 66 (51) and then for *Streptomyces coelicolor* A3(2) (52). A suicide vector was used to replace and link the ends of this linear chromosome. No differences in viability or segregation were found for the circular derivative so constructed (52). Strong noncovalent binding of the chromosomal ends, circle-linear oscillations, and pseudolinkage caused by repeats have been offered to explain the phenomena of circular genetic maps obtained for linear physical structures. One linear (2.2-Mb) and one circular (3.0-Mb) chromosome in *Agrobacterium tumefaciens* (1) shows another possible genome structure.

Physical maps: resolution, number of markers, and comparison with genetic maps. Most of the physical maps carry genetic markers localized mainly by hybridization. The resolution of these maps, defined as the size of a minimal element of the map (MRF or clone in encyclopedia) assigned to a certain gene, may vary from one to several hundred kilobases. Genes carrying Tn insertions marked by rare RSs can be mapped with much greater resolution. In this case, the gene is represented not by a positive fragment but by an additional site splitting it, which can be located with the accuracy of the electrophoresis used. Encyclopedias generally have much higher resolution than PFGE-generated maps, especially when restriction maps of their elements are constructed. Mapping cosmids with λ terminase simplifies this task (70).

The joint efforts of scientists in France, Germany, and the United States produced a combined physical-genetic map of *E. coli*. A total of 495 sequenced contigs were assigned to nearly 8,000 restriction fragments ordered in the Kohara map (58). The map itself was corrected in the areas where sequencing data were available, which already cover 50% of the genome. The Kohara map turned out to be highly accurate, although loss of closely spaced sites resulted in underestimating 5 to 15% of them (19). A total of 1,356 genes from the last edition of Bachmann's map were supplemented by 361 new ones, mainly derived from Rudd's EcoGene6 (74), and grouped by functions (72). Nearly 200 new genes were rescued from "scratches" (pieces of open reading frames in already screened sequences) by applying more-powerful gene search algorithms (8).

A total of 139 nonallelic auxotrophic mutations were assigned to cosmids of an encyclopedia of *H. volcanii* (20) by cosmid transformation of the mutant strains. A total of 54 loci of tRNA, 2 IS elements, and 12 cloned genes were located with high resolution on the *Mlu*I-digested cosmid miniset containing about 900 mapped sites.

About 70 probes were used to localize 200 genes identified in various *Rhodobacter* species on a cosmid miniset of *R. capsulatus* with an accuracy of 1 to 5 kb. The total number of RSs mapped on its genome exceeds 3,000. Sixty-three of 79 known genes were located on the map of *B. japonicum* (48). Twenty-eight new markers were added to the map of *N. gonorrhoeae* (27) on which about 60 genes had been located already. More than a hundred Tn elements in or near specific developmental genes were positioned on the 9.5-Mb chromosome of *M. xanthus* (14).

Clustering of genes related to certain functions, like the 46-kb "photosynthetic cluster" of *Rhodobacter* species, was shown for many microorganisms by conventional cloning or genetic mapping. It was expected that physical mapping of

genomes would reveal more such cases. However, many cases in which functionally linked genes are scattered all over the genome have been found instead.

Rearrangements and conservation of the *Ceu* skeleton. Sequencing data were used to establish phylogenetic relations among strains and to build evolutionary trees for individual genes. However, these trees for different genes have differences that, in some cases, may be attributed to recent acquisition of very remote foreign genes (horizontal transfer) (56). Even without horizontal transfer, recombinational exchange of gene blocks among closely related strains, similar to that described for lambdoid phages, can produce a mosaic structure of the genome. Evolutionary relations for genes within different blocks would be different and therefore a treelike scheme of strain relations would not be the best one.

Certain factors conserve gene order in chromosomes. Tn*10-plac-lacZ* inserted at various distances from the origin of the *E. coli* chromosome revealed a threefold expression gradient (61). Adaptation of gene regulation to this effect, caused by a higher copy number of DNA regions adjacent to the origin, can be one of such factors. Various inversions in the *E. coli* genome reduce its fitness (38), which can be restored by spontaneous reinversions.

Indirect evidence of a stable three-dimensional structure of the *Salmonella* chromosome resulted from a study of the distribution of the ends of rearrangements induced by Tn*10* (76). This transposon induced inversions that are clustered. The distance between the first end, fixed at the site of initial integration, and the second end is a multiple of 100 kb, the possible size of a conserved chromosomal domain. To discover more details of the conservation of elements of the genome among different bacteria, numerous comparisons of the physical maps were made. This task is difficult to fulfill by comparing remote groups of bacteria, because different markers are available for different groups. It is hard to align 120 developmental genes mapped in the *Myxococcus* genome with 200 genes mapped in the *Rhodobacter* genome that are mainly related to photosynthesis and nitrogen fixation. The problem of such comparisons can be solved if mapping is done with conserved PCR primers already described for *recA*, *rrl*, and genes of macromolecular biosynthesis operons and for some other major housekeeping genes. In a few cases, when such comparisons among distantly related organisms were performed, there was no visible conservation in the positions of any genome elements, possibly because of a fragmentary database. However, this comparison between closely related strains or species has revealed important conserved features.

We noticed that all organisms analyzed from this standpoint cover a broad spectrum of genome structure conservation. At one extreme, including *E. coli* and *Salmonella*, *Streptomyces*, *Neisseria*, *Lactococcus*, *Clostridium*, and *Mycoplasma* spp., almost absolute conservation of gene order can be traced by comparing different strains, species, and sometimes even different genera. At the other extreme, we find *Bacillus*, *Rhodobacter*, *Helicobacter* and perhaps *Leptospira* spp. and several cyanobacteria. Here, the gene order for different strains of the same species is highly rearranged, and no long-range conservation of the genetic map can be observed among members of the same genera.

By comparing *E. coli* and *S. typhimurium*, it has been shown that their gene order and populational structure are very strongly conserved. The world population of *E. coli*, more than a thousand isolates analyzed by multilocus protein electrophoresis (77), was called clonal because of the absence of traces of recombinational events. In studies establishing strain relations in *E. coli* and *Salmonella* spp. based on comparison of the gene sequences of proline permease (63) and many other proteins, a few cases of intragenic recombination were found. However, evidence for a mosaic structure of the *E. coli* chromosome was provided by Milkman (60), who used sequencing and multilocus restriction mapping of 15 regions of 72 strains.

High-resolution maps derived from encyclopedias are available for at least three strains of *E. coli* K-12: W3110 (44); BHB2600 (5) and 803 (43). PFGE-generated patterns for these strains differ in the positions of three or five *Not*I fragments of 22 (32a), typical for interstrain differences for various microorganisms. However, for the pair of W3110 and BHB2600, only 15 losses and 13 gains of *Eco*RI fragments (of 610 RSs for W3110) were observed (5). This number is close to the error of the experiment. The same high level of conservation of the restriction maps is demonstrated in the comparison of strains W3110 and 803 of *E. coli* (43). The large number of genome differences, shown for different strains of *E. coli*, ranging from almost a megabase inversion in the W3110 lineage (relative to most K-12 strains) to numerous small deletions, does not change the overall picture of this extreme stability. The same stability was found comparing different *Salmonella* species. Only half of the rare RSs were conserved at the same positions of the maps of *S. typhimurium* and *Salmonella paratyphi*, showing a relatively large evolutionary distance between these species. However, all 61 genes physically mapped by Tn*10* insertions in the chromosome of *S. paratyphi* are colinear in the two organisms (53). The level of divergence of nucleotide sequences of *E. coli* and *S. typhimurium* is 5 to 25% for different genes. It is impossible to see conservation in *Xba*I maps of the genomes of these organisms or in high-resolution maps of most areas. However, *Ceu*I restriction maps for several strains of *E. coli* and *S. typhimurium* (54) are almost identical, revealing nearly absolute conservation of the *rrn* operon skeleton for these two genera.

Other examples from the first group of bacteria are provided by maps of 10 strains of *Clostridium perfringens*, stapled by 11 gene probes and 17 anonymous cloned fragments (11). Colinearity of the maps made it possible to construct a consensus map where 10 variable areas were located. In the case of *S. lividans* and *S. coelicolor*, most linking clones were applicable to both organisms, the gene order was colinear, and PFGE-generated restriction maps resemble each other. These results are especially surprising, because this genus was considered to be highly recombinogenic, undergoing large chromosomal rearrangements (50).

Results of the comparison of four strains of *Bacillus cereus* and *Bacillus thuringiensis* are typical for organisms of the second group (12). Twenty-five probes were used to align the genomes of these strains, which are split into 10 to 16 intervals by mapped *Not*I sites. Even this low resolution is sufficient to reveal that the few patches of similarity in probe order are interrupted by numerous rearrangements, placing most of the genes as far as 1 to 2 Mb from their location in the other strains. Only two-thirds of the probes cloned from one of the *B. cereus* strains hybridize with all the others. In fact, two strains of *B. cereus* appear to be closer to *B. thuringiensis* than to the other *B. cereus* strains.

Genetic maps of two *Rhodobacter* species differ even in the number of chromosomes and *rrn* operons (33, 84). There is also no similarity in long-range gene order in any substantial segment of their chromosomes, though operon structure is similar for many genes studied. The *Eco*RV and *Bam*HI maps of 2-Mb segments of encyclopedias of three strains of *R. capsulatus* were compared by using 100 probes specific to the ends of linked cosmids. Conservation of 50 to 70% of RFs is sufficient for precise alignment of the fragments, confirmed by blot

hybridizations. The chosen area of the genome, uninterrupted in strain SB1003, is split into seven pieces in the St. Louis strain. This scrambled structure of the genome differs greatly from the structures seen in the first group of microorganisms.

APPLICATIONS AND FUTURE DEVELOPMENT

Gene mapping and repeated elements. PFGE-separated MRFs and clones making up an encyclopedia were used as hybridization templates in various mapping experiments. For example, a set of generalized probes complementary to parts of REPs (repetitive extragenic palindromic sequences) was used to probe the Kohara set (29). Ninety-four strongly positive and 23 ambiguously positive clones were found. A total of 112 such REP elements already exist in the database covering 38% of the *E. coli* genome. This points out that a database search can be as powerful as a direct experiment for a well-characterized organism.

Splitting the hybridization template into smaller pieces can help to reveal signals masked by higher DNA complexity in other types of hybridization. The miniset of 192 cosmids of *R. capsulatus* can be divided into 560 *Eco*RV fragments. Probing this gel blot with a fragment of the gene encoding glutamine permease revealed its precise location, but 20 additional weak signals were found in other *Eco*RV fragments. When the probe was split into two parts, one containing the conserved permease domain and the other containing its regulatory region, the group of weak signals was also divided into two sets.

Transcripts. The gene encyclopedia of *E. coli* was used to study its global regulatory network—a systematic description of genome expression at the RNA level. RNAs taken from cells grown under different conditions were hybridized with phage DNAs from the Kohara library (18). The hybridization targets were 20-kb phage inserts, which carried many genes, masking some differences in gene expression. Even so, differences of an order of magnitude were observed by comparing hybridization intensities produced by different RNAs. Total RNA induced by heat shock revealed 26 new heat shock-coexpressed genes, almost tripling the number of known heat shock genes. It was further demonstrated that cutting the DNA hybridization targets into smaller (gene-size) pieces raised the sensitivity 10 times (17).

Proteins. Systematic functional descriptions of bacterial genomes can be extended to the protein level by using gene encyclopedias as starting material. A derivative of pSC101 was used to express recloned inserts from three Kohara phages (75). An expression system with a very efficiently regulated T7 promoter was used. About a dozen genes were expressed from each insert in perfect agreement with the spots expected and identified on 2-D gels. Some new protein products were found and virtually total representation of the coding capacity was achieved even for genes totally repressed under standard physiological conditions. All *E. coli* proteins visible on 2-D gels can be linked to their genetic determinants by this approach.

CONCLUSIONS

It is both realistic and appropriate to anticipate the full characterization of bacterial chromosomes in terms of gene order, both short and long range, the global regulation of gene expression, and the complete elucidation of coding capacity in terms of proteins and nontranslated RNA. Achievement of these goals will require total sequencing, expression studies, and protein characterization.

Many paths have led to descriptions of the physical maps of bacterial chromosomes. If the goals of expression studies and protein characterization are included in the original formulation of strategy for a particular organism, it is clear that gene encyclopedias must be a component of the mapping scheme. We believe that cosmids such as Lorist 6 provide a good compromise with respect to insert size, being large enough to cover a bacterial chromosome with several hundred clones yet small enough to be mapped easily and displayed in gene-size fragments in a single gel lane. Mapped cosmids are also a good starting point for a total sequencing project, whether the sequencing is shotgun, is totally directed by PCR primers, or is a combination of the two. The cosmids provide constant feedback to the sequencers with respect to mapped genes, and they allow the chromosomal map to be developed logically.

Mapping projects have already revealed substantial plasticity in bacterial genomes. As sequencing projects proceed, close comparison of several closely related strains may define better the events that create the diversity seen at the level of the current physical maps. Finally, it must be pointed out that the vast amount of information that will flow from total sequencing projects will further increase the gaps between those researchers working on the few well-studied organisms and everyone else.

ACKNOWLEDGMENT

Work on the *Rhodobacter* genome in our laboratory is supported by grants from the Department of Energy and the Human Frontiers of Science program.

REFERENCES

1. **Allardet, S. A., C. S. Michaux, B. E. Jumas, L. Karayan, and M. Ramuz.** 1993. Presence of one linear and one circular chromosome in the *Agrobacterium tumefaciens* C58 genome. J. Bacteriol. **175:**7869–7874.
2. **Azevedo, V., E. Alvarez, E. Zumstein, G. Damiani, V. Sgaramella, S. D. Ehrlich, and P. Serror.** 1993. An ordered collection of *Bacillus subtilis* DNA segments cloned in yeast artificial chromosomes. Proc. Natl. Acad. Sci. USA **90:**6047–6051.
3. **Azevedo, V., A. Sorokin, S. D. Ehrlich, and P. Serror.** 1993. The transcriptional organization of the *Bacillus subtilis* 168 chromosome region between the *spoVA* and *serA* genetic loci. Mol. Microbiol. **10:**397–405.
4. **Bautsch, W.** 1988. Rapid physical mapping of the *Mycoplasma mobile* genome by two-dimensional field inversion gel electrophoresis techniques. Nucleic Acids Res. **16:**11461–11467.
5. **Birkenbihl, R. P., and W. Vielmetter.** 1989. Cosmid-derived map of *E. coli* strain BHB2600 in comparison to the map of strain W3110. Nucleic Acids Res. **17:**5057–5069.
6. **Blattner, F. R., V. Burland, G. Plunkett, H. J. Sofia, and D. L. Daniels.** 1993. Analysis of the *Escherichia coli* genome. IV. DNA sequence of the region from 89.2 to 92.8 minutes. Nucleic Acids Res. **21:**5408–5417.
7. **Borges, K. M., and P. L. Bergquist.** 1993. Genomic restriction map of the extremely thermophilic bacterium *Thermus thermophilus* HB8. J. Bacteriol. **175:**103–110.
8. **Borodovsky, M., E. V. Koonin, and K. E. Rudd.** 1994. New genes in old sequence: strategy for finding genes in the bacterial genome. Abstracts of a Conference on Genome Mapping and Sequencing, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
9. **Branscomb, E., T. Slezak, R. Pae, D. Galas, A. V. Carrano, and M. Waterman.** 1990. Optimizing restriction fragment fingerprinting methods for ordering large genomic libraries. Genomics **8:**351–366.
10. **Bukanov, N. O., and D. E. Berg.** 1994. Ordered cosmid library and high-resolution physical-genetic map of *Helicobacter pylori* strain NCTC11638. Mol. Microbiol. **11:**509–523.
11. **Canard, B., I. B. Saint Girons, and S. T. Cole.** 1992. Genomic diversity and organization of virulence genes in the pathogenic anaerobe *Clostridium perfringens*. Mol. Microbiol. **6:**1421–1429.
12. **Carlson, C. R., and A. B. Kolsto.** 1993. A complete physical map of a *Bacillus thuringiensis* chromosome. J. Bacteriol. **175:**1053–1060.
13. **Charlebois, R. L., L. C. Schalkwyk, J. D. Hofman, and W. F. Doolittle.** 1991. Detailed physical map and set of overlapping clones covering the genome of the archaebacterium *Haloferax volcanii* DS2. J. Mol. Biol. **222:**509–524.
14. **Chen, H. W., A. Kuspa, I. M. Keseler, and L. J. Shimkets.** 1991. Physical map of the *Myxococcus xanthus* chromosome. J. Bacteriol. **173:**2109–2115.
15. **Chen, X., and W. R. Widger.** 1993. Physical genome map of the unicellular cyanobacterium *Synechococcus* sp. strain PCC 7002. J. Bacteriol. **175:**5106–5116.
16. **Cheng, H.-P., and T. G. Lessie.** 1994. Multiple replicons constituting the

genome of *Pseudomonas cepacia* 17616. J. Bacteriol. **176:**4034–4042.
17. **Chuang, S. E., V. Burland, G. Plunkett, D. L. Daniels, and F. R. Blattner.** 1993. Sequence analysis of four new heat-shock genes constituting the *hslTS*/*ibpAB* and *hslVU* operons in *Escherichia coli*. Gene **134:**1–6.
18. **Chuang, S. E., D. L. Daniels, and F. R. Blattner.** 1993. Global regulation of gene expression in *Escherichia coli*. J. Bacteriol. **175:**2026–2036.
19. **Churchill, G. A., D. L. Daniels, and M. S. Waterman.** 1990. The distribution of restriction enzyme sites in *Escherichia coli*. Nucleic Acids Res. **18:**589–597.
20. **Cohen, A., W. L. Lam, R. L. Charlebois, W. F. Doolittle, and L. C. Schalkwyk.** 1992. Localizing genes on the map of the genome of *Haloferax volcanii*, one of the Archaea. Proc. Natl. Acad. Sci. USA **89:**1602–1606.
21. **Cole, S. T., and I. Saint Girons.** 1994. Bacterial genomics. FEMS Microbiol. Rev. **14:**139–160.
22. **Colman, S. D., P. C. Hu, W. Litaker, and K. F. Bott.** 1990. A physical map of the *Mycoplasma genitalium* genome. Mol. Microbiol. **4:**683–687.
23. **Coulson, A., J. Sulston, S. Brenner, and J. Karn.** 1986. Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA **83:**7821–7825.
24. **Craig, A. G., D. Nizetic, J. D. Hoheisel, G. Zehetner, and H. Lehrach.** 1990. Ordering of cosmid clones covering the herpes simplex virus type I (HSV-I) genome: a test case for fingerprinting by hybridisation. Nucleic Acids Res. **18:**2653–2660.
25. **Daniels, D. L.** 1990. Constructing encyclopedias of genomes, p. 43–52. *In* K. Drlica and M. Riley (ed.), The bacterial chromosome. American Society for Microbiology, Washington, D.C.
26. **Deckers, H. M., and G. Voordouw.** 1994. Identification of a large family of genes for putative chemoreceptor proteins in an ordered library of the *Desulfovibrio vulgaris* Hildenborough genome. J. Bacteriol. **176:**351–358.
27. **Dempsey, J. A., and J. G. Cannon.** 1994. Locations of genetic markers on the physical map of the chromosome of *Neisseria gonorrhoeae* FA1090. J. Bacteriol. **176:**2055–2060.
28. **Dempsey, J. A., W. Litaker, A. Madhure, T. L. Snodgrass, and J. G. Cannon.** 1991. Physical map of the chromosome of *Neisseria gonorrhoeae* FA1090 with locations of genetic markers, including *opa* and *pil* genes. J. Bacteriol. **173:** 5476–5486.
29. **Dimri, G. P., K. E. Rudd, M. K. Morgan, H. Bayat, and G. F. Ames.** 1992. Physical mapping of repetitive extragenic palindromic sequences in *Escherichia coli* and phylogenetic distribution among *Escherichia coli* strains and other enteric bacteria. J. Bacteriol. **174:**4583–4593.
30. **Eiglmeier, K., N. Honore, S. A. Woods, B. Caudron, and S. T. Cole.** 1993. Use of an ordered cosmid library to deduce the genomic organization of *Mycobacterium leprae*. Mol. Microbiol. **7:**197–206.
31. **Ely, B., T. W. Ely, C. J. Gerardot, and A. Dingwall.** 1990. Circularity of the *Caulobacter crescentus* chromosome determined by pulsed-field gel electrophoresis. J. Bacteriol. **172:**1262–1266.
32. **Ferdows, M. S., and A. G. Barbour.** 1989. Megabase-sized linear DNA in the bacterium *Borrelia burgdorferi*, the Lyme disease agent. Proc. Natl. Acad. Sci. USA **86:**5969–5973.
32a. **Fonstein, M.** Submitted for publication.
33. **Fonstein, M., and R. Haselkorn.** 1993. Chromosomal structure of *Rhodobacter capsulatus* strain SB1003: cosmid encyclopedia and high-resolution physical and genetic map. Proc. Natl. Acad. Sci. USA **90:**2522–2526.
34. **Fonstein, M., S. Zheng, and R. Haselkorn.** 1992. Physical map of the genome of *Rhodobacter capsulatus* SB 1003. J. Bacteriol. **174:**4070–4077.
35. **Fu, Y. X., W. E. Timberlake, and J. Arnold.** 1992. On the design of genome mapping experiments using short synthetic oligonucleotides. Biometrics **48:** 337–359.
36. **Gibson, T. J., A. Rosenthal, and R. H. Waterston.** 1987. Lorist 6, a cosmid vector with *Bam*HI, *Not*I, *Sca*I and *Hin*dIII cloning sites and altered neomycin phosphotransferase gene expression. Gene **53:**283–286.
37. **Glaser, P., F. Kunst, M. Arnaud, M.-P. Coudart, W. Gonzales, M.-F. Hullo, M. Ionescu, B. Lubochinsky, L. Marcelino, I. Moszer, E. Presecan, M. Santana, E. Schneider, J. Schweizer, A. Vertes, G. Rapoport, and A. Danchin.** 1993. *Bacillus subtilis* genome project: cloning and sequencing of the 97-kb region from 325° to 333°. Mol. Microbiol. **10:**371–384.
38. **Hill, C. W., and J. A. Gray.** 1988. Effects of chromosomal inversion on cell fitness in *Escherichia coli* K-12. Genetics **119:**771–778.
39. **Hoheisel, J. D., E. Maier, R. Mott, L. McCarthy, A. V. Grigoriev, L. C. Schalkwyk, D. Nizetic, F. Francis, and H. Lehrach.** 1993. High resolution cosmid and P1 maps spanning the 14-Mb genome of the fission yeast *S. pombe*. Cell **73:**109–120.
40. **Honeycutt, R. J., M. McClelland, and B. W. Sobral.** 1993. Physical map of the genome of *Rhizobium meliloti* 1021. J. Bacteriol. **175:**6945–6952.
41. **Honore, N., S. Bergh, S. Chanteau, P. F. Doucet, K. Eiglmeier, T. Garnier, C. Georges, P. Launois, T. Limpaiboon, S. Newton, et al.** 1993. Nucleotide sequence of the first cosmid from the *Mycobacterium leprae* genome project: structure and function of the Rif-Str regions. Mol. Microbiol. **7:**207–214.
42. **Itaya, M., and T. Tanaka.** 1991. Complete physical map of the *Bacillus subtilis* 168 chromosome constructed by a gene-directed mutagenesis method. J. Mol. Biol. **220:**631–648.
43. **Knott, V., D. J. Blake, and G. G. Brownlee.** 1989. Completion of the detailed restriction map of the *E. coli* genome by the isolation of overlapping cosmid clones. Nucleic Acids Res. **17:**5901–5912.
44. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50:**495–508.
45. **Koob, M., A. Burkiewicz, J. Kur, and W. Szybalski.** 1992. RecA-AC: single-site cleavage of plasmids and chromosomes at any predetermined restriction site. Nucleic Acids Res. **20:**5831–5836.
46. **Koob, M., and W. Szybalski.** 1990. Cleaving yeast and *Escherichia coli* genomes at a single site. Science **250:**271–273.
47. **Kroger, M., R. Wahl, and P. Rice.** 1993. Compilation of DNA sequences of *Escherichia coli* (update 1993). Nucleic Acids Res. **21:**2973–3000.
48. **Kundig, C., H. Hennecke, and M. Gottfert.** 1993. Correlated physical and genetic map of the *Bradyrhizobium japonicum* 110 genome. J. Bacteriol. **175:**613–622.
49. **Kuspa, A., D. Vollrath, Y. Cheng, and D. Kaiser.** 1989. Physical mapping of the *Myxococcus xanthus* genome by random cloning in yeast artificial chromosomes. Proc. Natl. Acad. Sci. USA **86:**8917–8921.
50. **Leblond, P., P. Demuyter, J. M. Simonet, and B. Decaris.** 1991. Genetic instability and associated genome plasticity in *Streptomyces ambofaciens*: pulsed-field gel electrophoresis evidence for large DNA alterations in a limited genomic region. J. Bacteriol. **173:**4229–4233.
51. **Leblond, P., M. Redenbach, and J. Cullum.** 1993. Physical map of the *Streptomyces lividans* 66 genome and comparison with that of the related strain *Streptomyces coelicolor* A3(2). J. Bacteriol. **175:**3422–3429.
52. **Lin, Y.-S., H. M. Kieser, D. A. Hopwood, and C. W. Chen.** 1993. The chromosomal DNA of *Streptomyces lividans* 66 is linear. Mol. Microbiol. **10:**923–933.
53. **Liu, S. L., A. Hessel, H. Y. Cheng, and K. E. Sanderson.** 1994. The *Xba*I-*Bln*I-*Ceu*I genomic cleavage map of *Salmonella paratyphi* B. J. Bacteriol. **176:**1014–1024.
54. **Liu, S. L., A. Hessel, and K. E. Sanderson.** 1993. Genomic mapping with I-*Ceu* I, an intron-encoded endonuclease specific for genes for ribosomal RNA, in *Salmonella* spp., *Escherichia coli*, and other bacteria. Proc. Natl. Acad. Sci. USA **90:**6874–6878.
55. **Liu, S. L., and K. E. Sanderson.** 1992. A physical map of the *Salmonella typhimurium* LT2 genome made by using *Xba*I analysis. J. Bacteriol. **174:** 1662–1672.
56. **Mazodier, P., and J. Davies.** 1991. Gene transfer between distantly related bacteria. Annu. Rev. Genet. **25:**147–171.
57. **McClelland, M., R. Jones, Y. Patel, and M. Nelson.** 1987. Restriction endonucleases for pulsed field mapping of bacterial genomes. Nucleic Acids Res. **15:**5985–6005.
58. **Medigue, C., A. Viari, A. Henaut, and A. Danchin.** 1993. Colibri: a functional data base for the *Escherichia coli* genome. Microbiol. Rev. **57:**623–654.
59. **Meier, E. S., E. Maier, A. Ahmadi, J. Curtis, and H. Lehrach.** 1993. An automated approach to generating expressed sequence catalogues. Nature (London) **361:**375–376.
60. **Milkman, R., and M. M. Bridges.** 1993. Molecular evolution of the *Escherichia coli* chromosome. IV. Sequence comparisons. Genetics **133:**455–468.
61. **Miller, W. G., and R. W. Simons.** 1993. Chromosomal supercoiling in *Escherichia coli*. Mol. Microbiol. **10:**675–684.
62. **Mizukami, T., W. I. Chang, I. Garkavtsev, N. Kaplan, D. Lombardi, T. Matsumoto, O. Niwa, A. Kounosu, M. Yanagida, T. G. Marr, et al.** 1993. A 13 kb resolution cosmid map of the 14 Mb fission yeast genome by nonrandom sequence-tagged site mapping. Cell **73:**121–132.
63. **Nelson, K., and R. K. Selander.** 1992. Evolutionary genetics of the proline permease gene (*putP*) and the control region of the proline utilization operon in populations of *Salmonella* and *Escherichia coli*. J. Bacteriol. **174:** 6886–6895.
64. **Neumann, B., A. Pospiech, and H. U. Schairer.** 1992. Size and stability of the genomes of the myxobacteria *Stigmatella auriantica* and *Stigmatella erecta*. J. Bacteriol. **174:**6307–6310.
65. **Nishimura, A., K. Akiyama, Y. Kohara, and K. Horiuchi.** 1992. Correlation of a subset of the pLC plasmids to the physical map of *Escherichia coli* K-12. Microbiol. Rev. **56:**137–151.
66. **Noll, K. M.** 1989. Chromosome map of the thermophilic archaebacterium *Thermococcus celer*. J. Bacteriol. **171:**6720–6725.
67. **Olson, M. V., J. E. Dutchik, M. Y. Graham, G. M. Brodeur, C. Helms, M. Frank, M. MacCollin, R. Scheinman, and T. Frank.** 1986. Random-clone strategy for genomic restriction mapping in yeast. Proc. Natl. Acad. Sci. USA **83:**7826–7830.
68. **Peterson, S. N., N. Schramm, P. C. Hu, K. F. Bott, and C. A. Hutchison.** 1991. A random sequencing approach for placing markers on the physical map of *Mycoplasma genitalium*. Nucleic Acids Res. **19:**6027–6031.
69. **Poddar, S. K., and M. McClelland.** 1992. Nonspecific primer and PCR-generated hybridization probes for physical ordering of large restriction fragments in the complex genome of *S. aureus*. Biochem. Biophys. Res. Commun. **185:**1055–1061.
70. **Rackwitz, H. R., G. Zehetner, H. Murialdo, H. Delius, J. H. Chai, A. Poustka, A. Frischauf, and H. Lehrach.** 1985. Analysis of cosmids using linearization by phage lambda terminase. Gene **40:**259–266.

71. **Riley, M.** 1993. Functions of the gene products of *Escherichia coli*. Microbiol. Rev. **57:**862–952.
72. **Romling, U., D. Grothues, W. Bautsch, and B. Tummler.** 1989. A physical genome map of *Pseudomonas aeruginosa* PAO. EMBO J. **8:**4081–4089.
73. **Rudd, K. E., G. Bouffard, and W. Miller.** 1992. Computer analysis of *Escherichia coli* restriction maps, p. 1–38. *In* K. Davies and S. Tilghman (ed.), Genome analysis. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
74. **Sankar, P., M. E. Hutton, R. A. VanBogelen, R. L. Clark, and F. C. Neidhardt.** 1993. Expression analysis of cloned chromosomal segments of *Escherichia coli*. J. Bacteriol. **175:**5145–5152. (Erratum, **176:**939, 1994.)
75. **Schmid, M. B.** 1993. Clustering of the endpoints of the inversions induced by Tn*10* in *Salmonella typhimurium*. Second E. coli Genome International Meeting, Madison, Wis. Supplement page.
76. **Schwartz, D. C., and C. R. Cantor.** 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell **37:**67–75.
77. **Selander, R. K., D. A. Caugant, H. Ochman, J. M. Musser, M. N. Gilmour, and T. S. Whittam.** 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl. Environ. Microbiol. **51:**873–884.
78. **Shizuya, H., B. Birren, U. J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, and M. Simon.** 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc. Natl. Acad. Sci. USA **89:**8794–8797.
79. **Smith, C. L., J. G. Econome, A. Schutt, S. Klco, and C. R. Cantor.** 1987. A physical map of the *Escherichia coli* K12 genome. Science **236:**1448–1453.
80. **Smith, C. L., and R. D. Kolodner.** 1988. Mapping of *Escherichia coli* chromosomal Tn*5* and F insertions by pulsed field gel electrophoresis. Genetics **119:**227–236.
81. **Sorokin, A., E. Zumstein, V. Azevedo, S. D. Ehrlich, and P. Serror.** 1993. The organization of the *Bacillus subtilis* 168 chromosome region between the *spoVA* and *serA* genetic loci, based on sequence data. Mol. Microbiol. **10:** 385–395.
82. **Strobel, S. A., and P. B. Dervan.** 1991. Single-site enzymatic cleavage of yeast genomic DNA mediated by triple helix formation. Nature (London) **350:** 172–174.
83. **Suwanto, A., and S. Kaplan.** 1989. Physical and genetic mapping of the *Rhodobacter sphaeroides* 2.4.1 genome: presence of two unique circular chromosomes. J. Bacteriol. **171:**5850–5859.
84. **Tigges, E., and F. C. Minion.** 1994. Physical map of the genome of *Acholeplasma oculi* ISM1499 and construction of a Tn*4001* derivative for macrorestriction chromosomal mapping. J. Bacteriol. **176:**1180–1183.
85. **Ventra, L., and A. S. Weiss.** 1989. Transposon-mediated restriction mapping of the *Bacillus subtilis* chromosome. Gene **78:**29–36.
86. **Wang, Y., and R. Wu.** 1993. A new method for specific cleavage of megabase-size chromosomal DNA by λ-terminase. Nucleic Acids Res. **21:**2143–2147.
87. **Wenzel, R., and R. Herrmann.** 1989. Cloning of the complete *Mycoplasma pneumoniae* genome. Nucleic Acids Res. **17:**7029–7043.
88. **Young, D. B., and S. T. Cole.** 1993. Leprosy, tuberculosis, and the new genetics. J. Bacteriol. **175:**1–6.
89. **Zuerner, R. L.** 1991. Physical map of chromosomal and plasmid DNA comprising the genome of *Leptospira interrogans*. Nucleic Acids Res. **19:**4857–4860.